



UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FCFM

FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS



Práctica 1: Análisis exploratorio

Aprendizaje Automático

Nombre: Sergio Andrés Elizondo Rodríguez

Grupo 003

Maestro: José Anastacio Hernández Saldaña

Sábado 29 de junio del 2024

Análisis exploratorio de salarios en la UANL

El **análisis exploratorio** es una fase crucial en un proyecto de ciencia de datos, ya que nos permite tener un panorama inicial del conjunto a estudiar, descubriendo patrones, anomalías y relaciones entre las variables.

En esta práctica, se utiliza la información sobre salarios por dependencia de la Universidad Autónoma de Nuevo León (UANL) para realizar un análisis exploratorio completo.

Objetivos del análisis

- Identificar y describir las entidades del conjunto de datos.
- Calcular estadísticas descriptivas básicas (mínimo, máximo, promedio, desviación típica, etc.) para cada entidad.
- Realizar agrupaciones de las diferentes entidades y obtener estadísticas descriptivas para cada grupo.
- Crear gráficos (histogramas, de pastel, etc.) que representen las estadísticas obtenidas, incluyendo las de las agrupaciones.
- Realizar pruebas de análisis de varianza (ANOVA) para determinar si existen diferencias significativas entre grupos (o bien, pruebas de Kruskal-Wallis y Tukey en caso de que las muestras no sean normales).

Identificación de entidades

En este conjunto de datos se han identificado varias **entidades** clave que pueden ser objeto de estudio:

- **Empleados:** Individuos que reciben sueldos en la UANL. Cada registro del conjunto contiene información relacionada con el empleado, incluyendo su sueldo neto y la dependencia a la que pertenece.

- **Dependencias:** Áreas, departamentos o unidades dentro de la UANL en las que trabajan los empleados. El análisis de las dependencias permite, por ejemplo, comprender la distribución y variación de sueldos.
- **Tipos de empleados:** Categorías o roles dentro de la universidad. Es útil para entender cómo se distribuyen los sueldos entre los distintos tipos de empleo.

Análisis descriptivo de entidades

A continuación, se presentan los hallazgos más significativos de las entidades de acuerdo con el análisis descriptivo realizado:

Empleados

Al analizar los sueldos netos de los empleados, se observaron las siguientes estadísticas descriptivas:

Estadística	Valor
Cantidad de registros	636,201 empleados
Sueldo neto promedio	\$14,241.68 MXN
Desviación típica del sueldo neto	\$9,578.44 MXN
Sueldo neto mínimo	\$175.41 MXN
Primer cuartil (25%)	\$8,007.66 MXN
Mediana (50%)	\$11,426.50 MXN
Tercer cuartil (75%)	\$17,654.63 MXN
Sueldo neto máximo	\$147,051.59 MXN

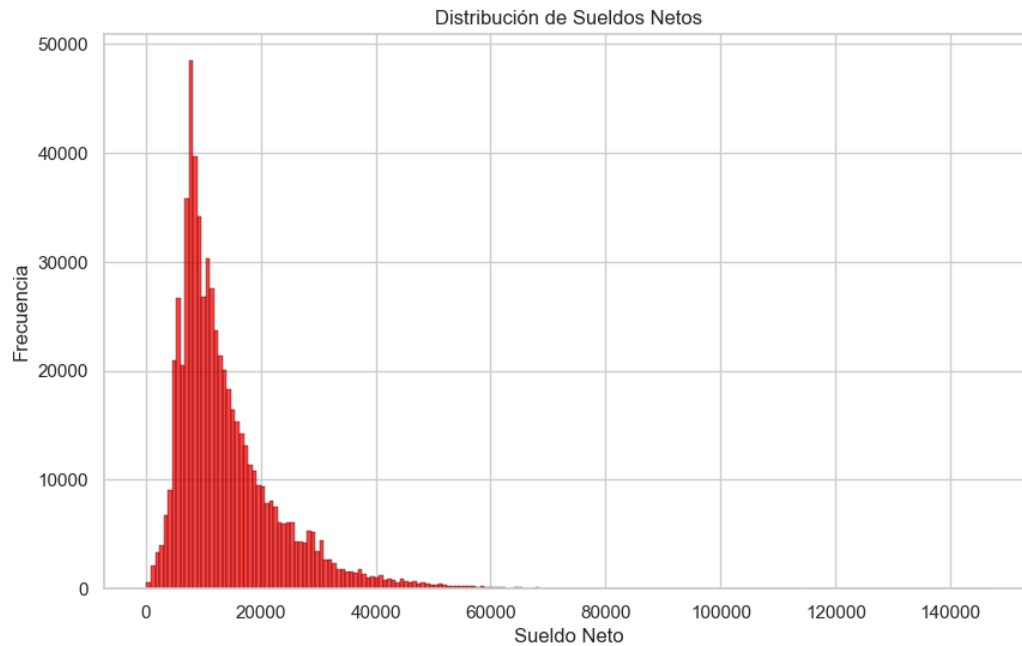


Ilustración 1: Distribución de sueldos netos.

La desviación típica y la diferencia considerable entre el sueldo mínimo y máximo sugieren una desigualdad significativa en la distribución de sueldos dentro de la UANL.

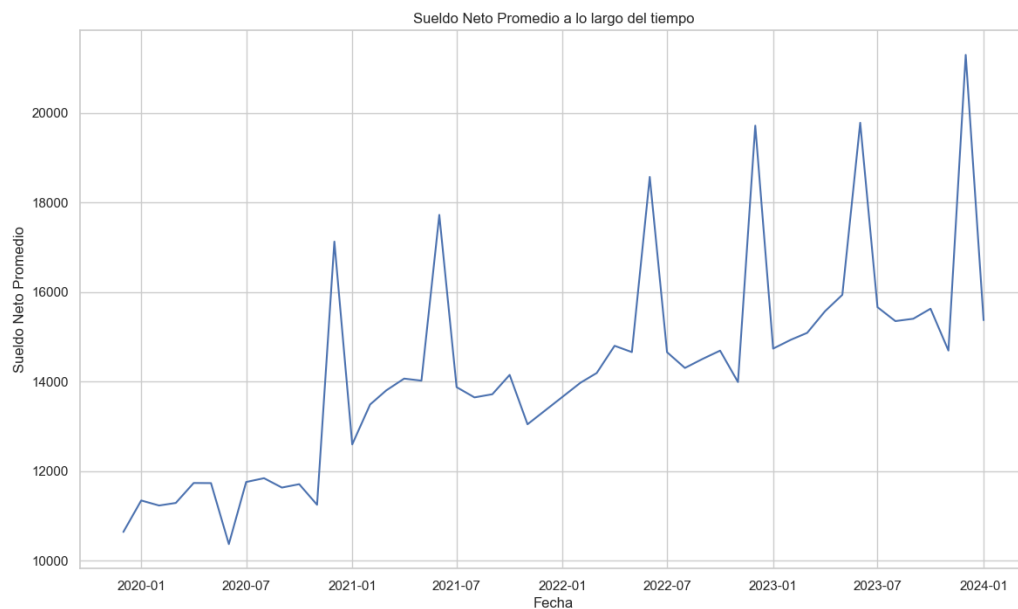


Ilustración 2: Sueldo neto promedio a lo largo del tiempo.

Dependencias

Al analizar los sueldos por dependencia, se obtuvieron algunos hallazgos importantes:

- La **Dirección de Unidades Académicas** tiene el sueldo promedio más alto (\$48,706.90 MXN).
- La **Dirección de Planeación Estratégica en Innovación** tiene el sueldo promedio más bajo (\$8,792.03 MXN). Además, esta dependencia tiene la menor desviación típica (\$2,051.25 MXN), lo cual indica una baja dispersión de sueldos.

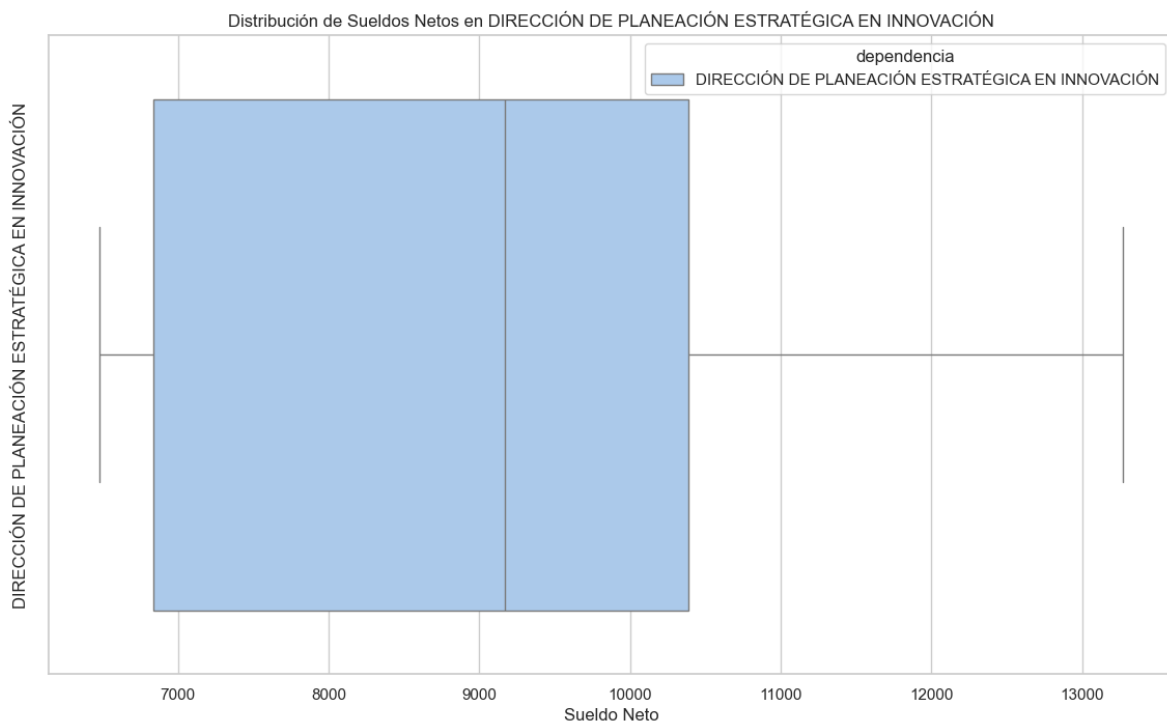


Ilustración 3: Variabilidad de sueldos en la Dirección de Planeación Estratégica en Innovación.

- La **Rectoría** tiene la mayor dispersión de sueldos, con una desviación típica de \$27,316.33 MXN, lo cual indica una mayor desigualdad en los sueldos de dicha dependencia.

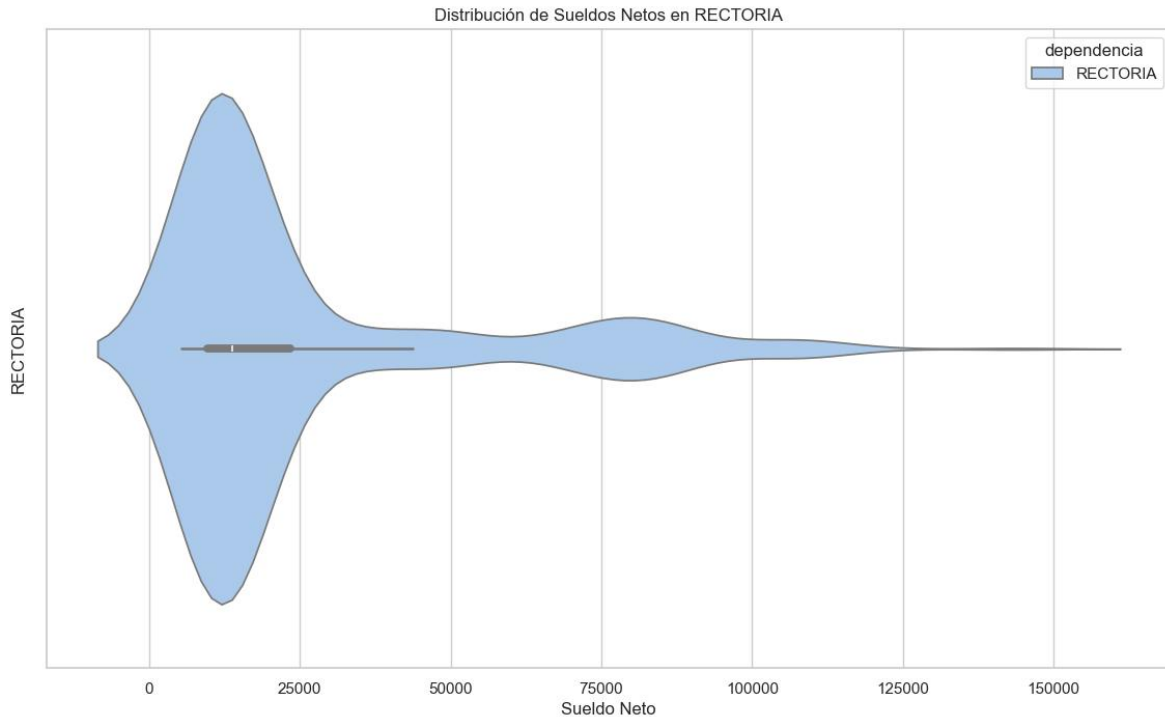


Ilustración 5: Variabilidad de sueldos netos en la Rectoría.

- La **Dirección del Programa de Fútbol Americano** tiene el primer percentil de sueldos más bajo (\$6,415.04); esto indica que el 25% de los empleados de esta dirección gana menos de dicha cantidad (la cual se encuentra incluso por debajo del salario mínimo).
- La **Dirección de Investigación Educativa** tiene el tercer percentil de sueldos más alto (\$62,321.93), lo cual indica que el 25% de los empleados de esta dependencia gana más de dicha cantidad.

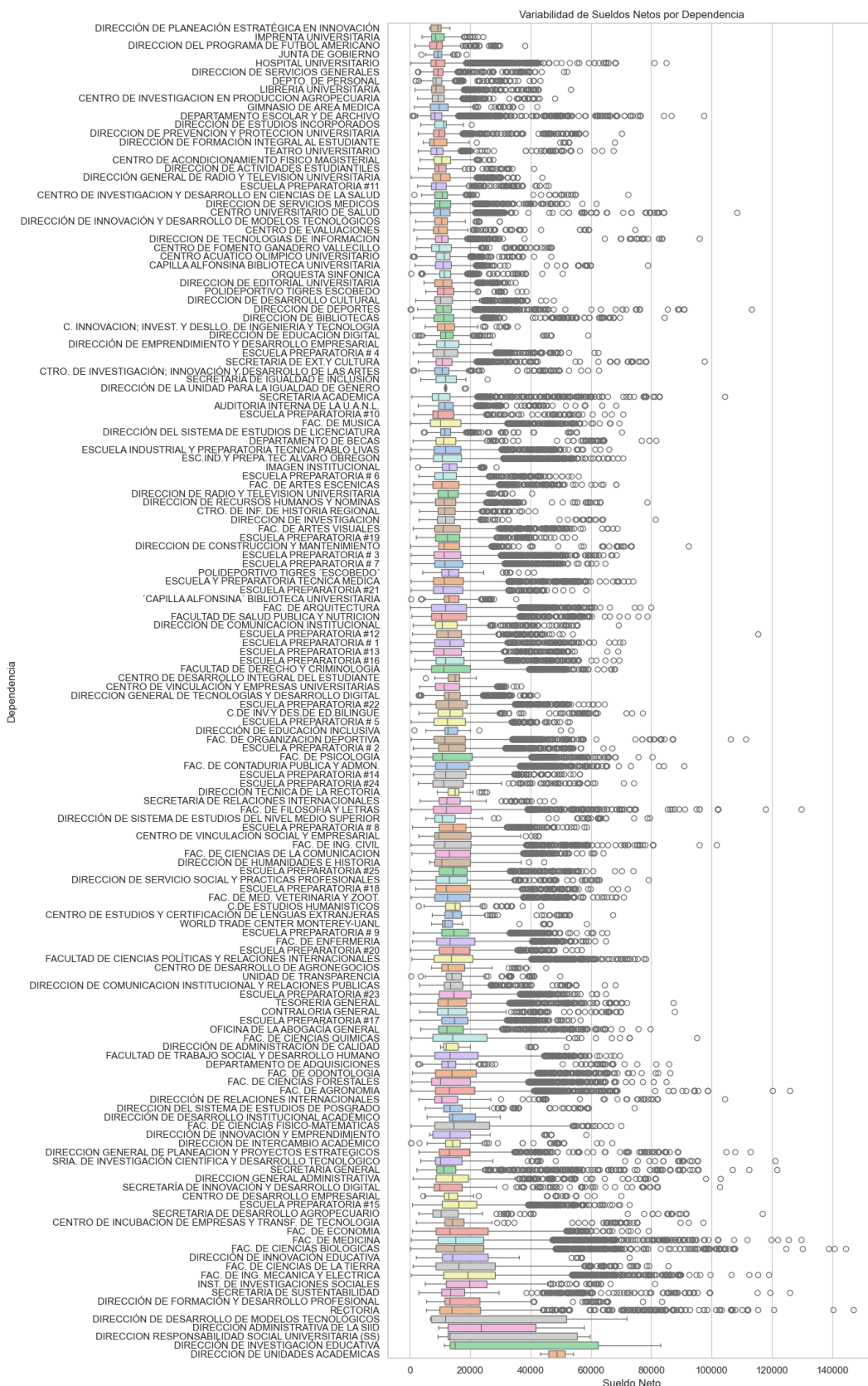


Ilustración 6: Variabilidad de sueldos netos por dependencia.

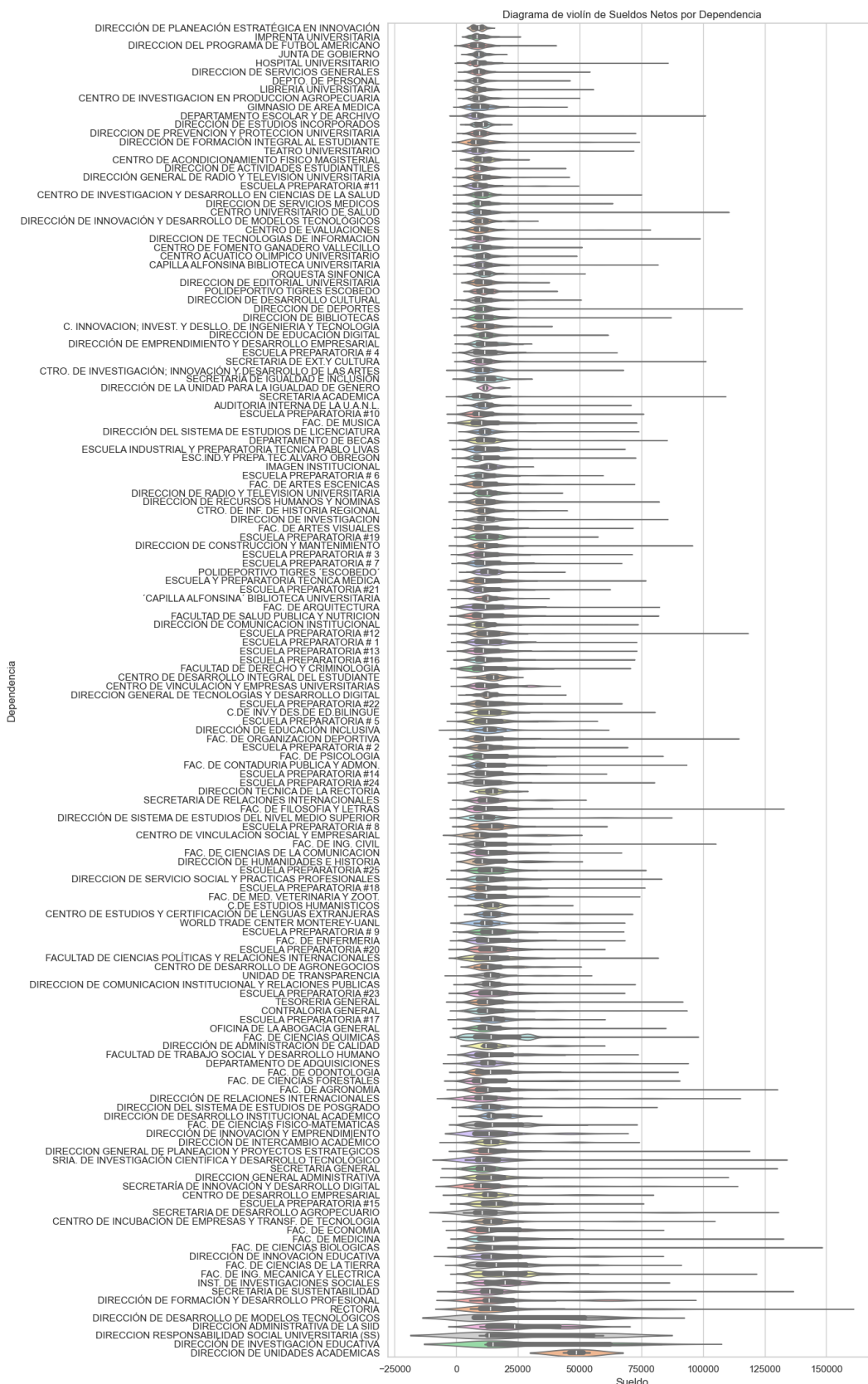


Ilustración 7: Diagrama de violín de sueldos netos por dependencia.

- La dependencia con más empleados es el **Hospital Universitario**, con 105,549 personas recibiendo salario.
- La dependencia con menos empleados es la **Dirección de Unidades Académicas**, la cual tiene tan solo dos registros de nómina.

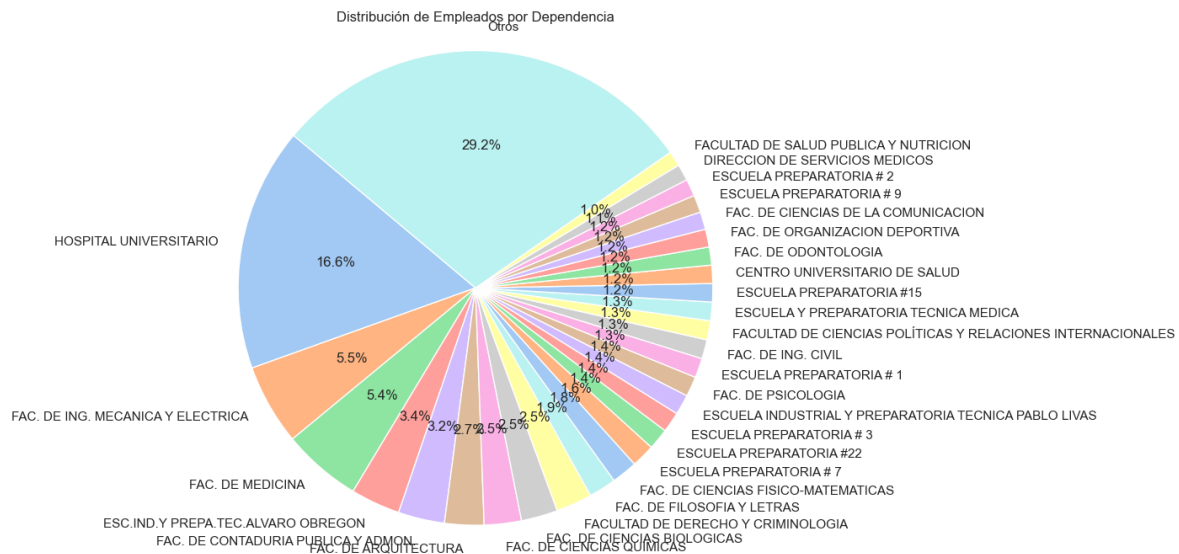


Ilustración 8: Distribución de empleados por dependencia.

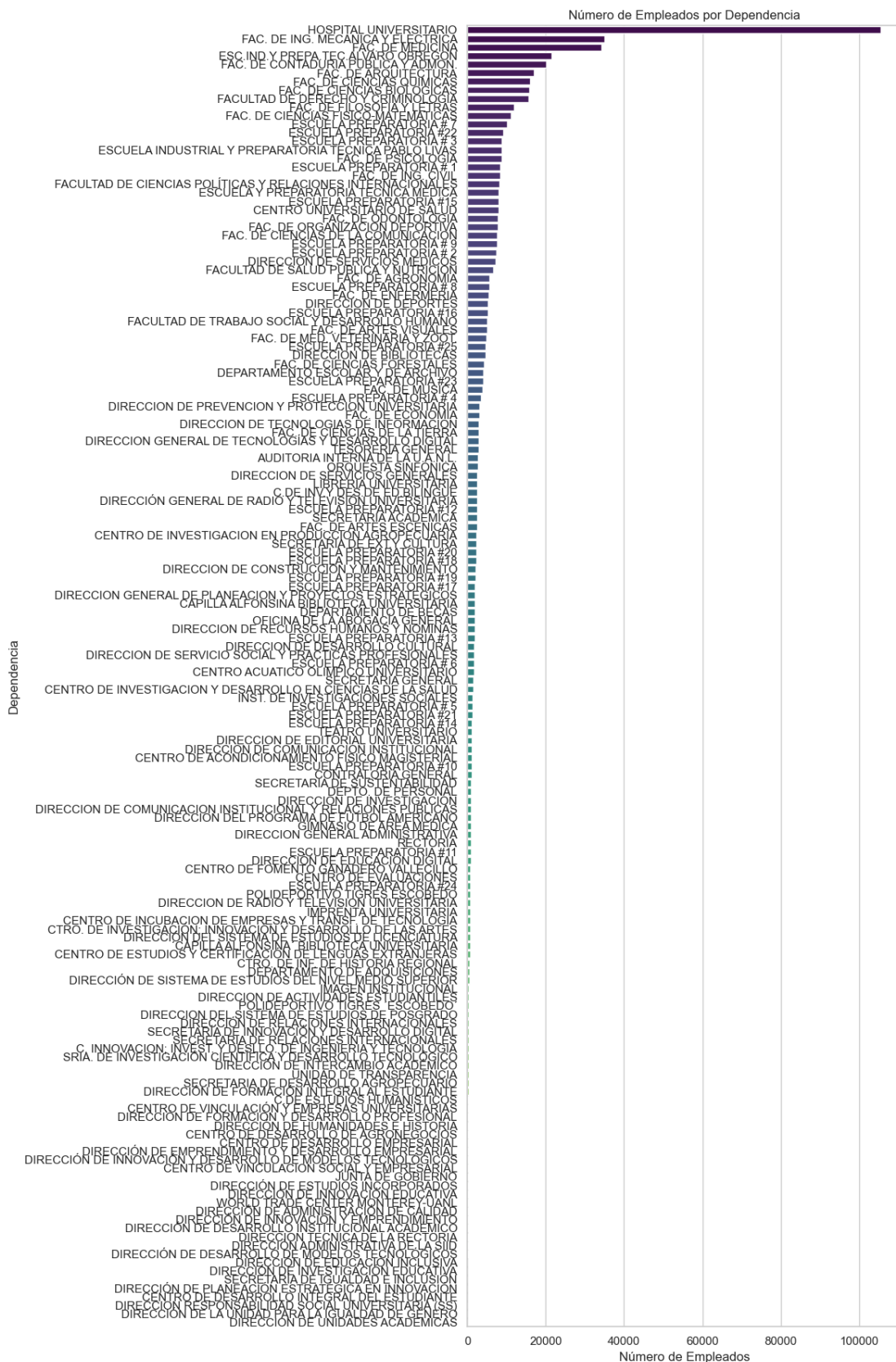


Ilustración 9: Conteo de empleados por dependencia

Tipos de Empleado

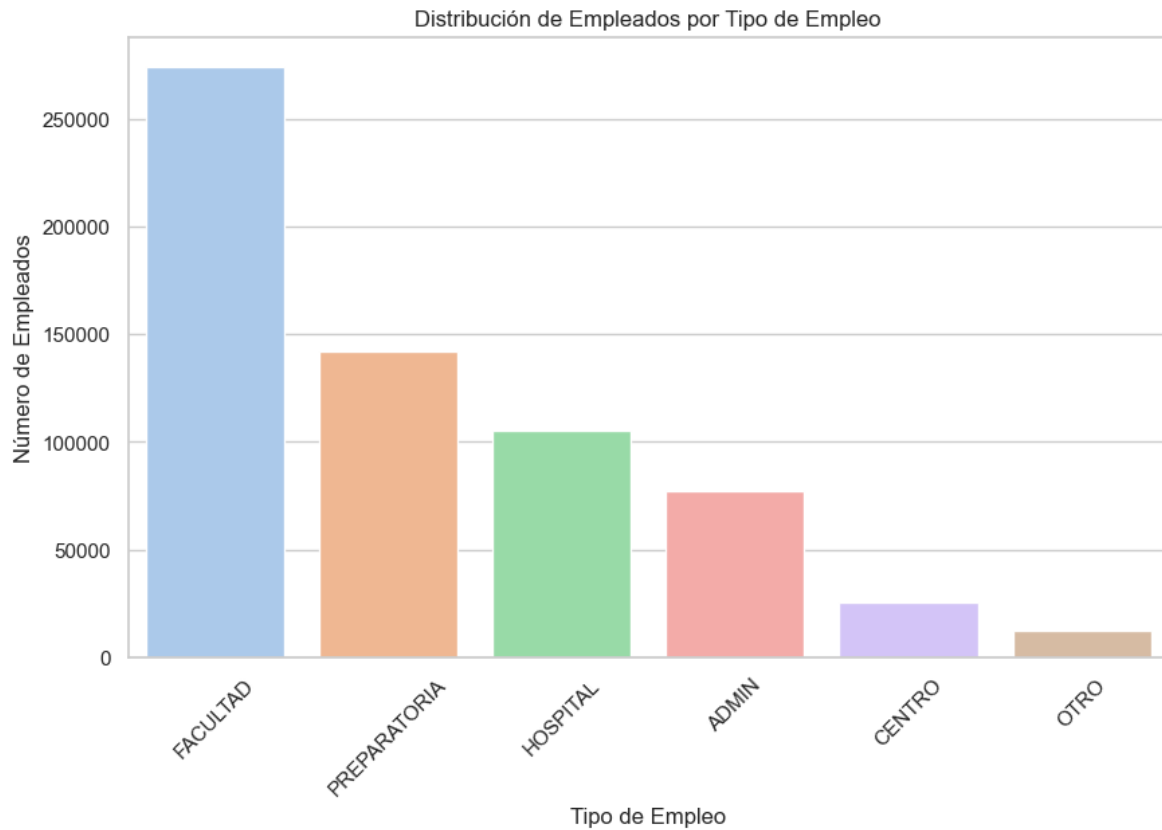


Ilustración 10: Distribución de empleados por tipo de empleo.

- Los empleados de **facultad** tienen el sueldo neto promedio más alto (16,515 MXN), seguido por los de **preparatoria** (14,281 MXN) y los **administrativos** (13,304 MXN).
- Los empleados del **hospital** tienen el sueldo neto promedio más bajo (9,631 MXN).
- La desviación típica es más alta para los empleados de **facultad** (\$11,037 MXN), lo cual sugiere una mayor variabilidad entre los sueldos de este grupo.
- La desviación típica más baja es la de los empleados del **hospital** (\$4,169 MXN), lo cual sugiere una mayor estabilidad en sueldos.

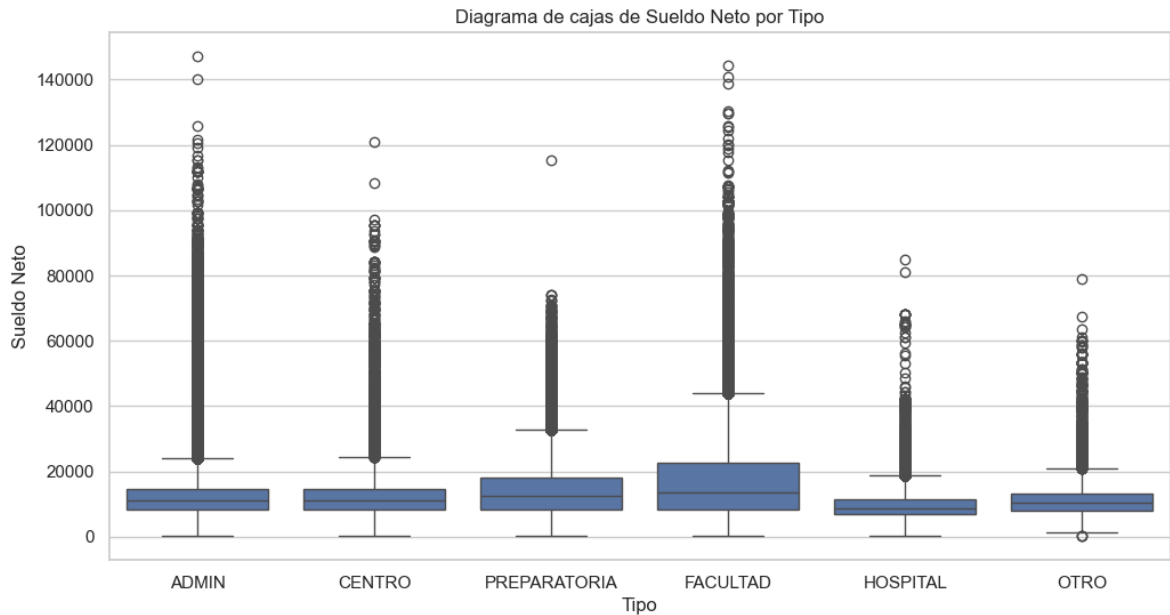


Ilustración 11: Diagrama de cajas de sueldo neto por tipo de empleo.

Análisis estadístico de sueldos

Pruebas de normalidad por dependencia

Se realizaron pruebas de normalidad para verificar si los sueldos netos en cada dependencia seguían una distribución normal. Para ello, se utilizó la **prueba de Kolmogórov-Smirnov (K-S)** con el fin de comparar la distribución de los sueldos con la distribución normal teórica, bajo la **hipótesis nula** de que la muestra proviene de dicha distribución.

Los resultados obtenidos mostraron que, para todas las dependencias, el valor del **estadístico K-S** fue **1.0** y el **valor p** fue **0.0**, lo cual indica que rechazamos la hipótesis nula en todos los casos. Esto sugiere que los sueldos netos no siguen una distribución normal en ninguna dependencia, por lo cual no es posible realizar una prueba de análisis de varianza (ANOVA) tradicional.

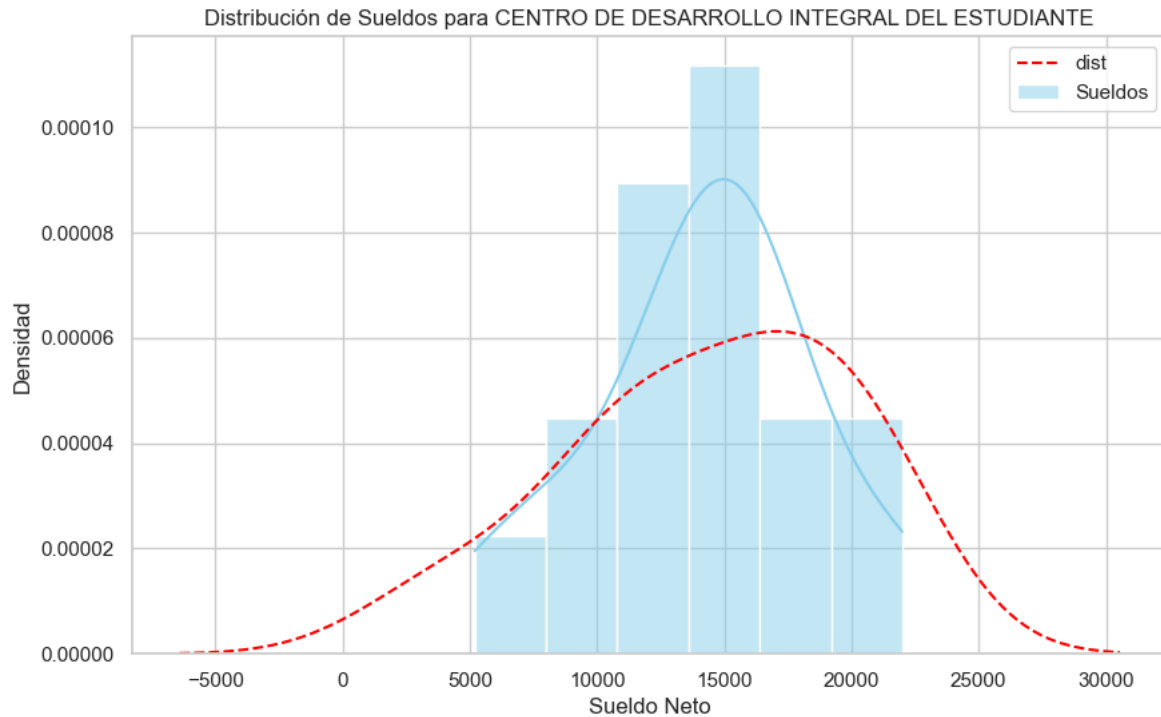


Ilustración 12: El Centro de Desarrollo Integral del Estudiante es el que presenta sueldos distribuidos más normalmente de acuerdo con la prueba de Shapiro-Wilks.

Prueba de Kruskal-Wallis

En su lugar, recurrimos a **Kruskal-Wallis**, una **prueba no paramétrica** para determinar si existen diferencias significativas entre las medianas de más de dos grupos independientes. En nuestro caso, aplicamos dicha prueba para comparar los sueldos netos entre las 6 dependencias con menor sueldo promedio:

- Dirección de Planeación Estratégica en Innovación
- Imprenta Universitaria
- Dirección del Programa de Fútbol Americano
- Junta de Gobierno
- Hospital Universitario
- Dirección de Servicios Generales

El valor del **estadístico de la prueba** fue **37.3504**, correspondiente a un **valor p** de aproximadamente **5.09×10^{-9}** , menor que el nivel de significancia de 0.05.

Por lo tanto, rechazamos la hipótesis nula de que todas las medianas son iguales, lo cual indica que al menos una de las dependencias tiene un sueldo neto medio significativamente diferente.

Prueba de Tukey (análisis post-hoc)

Para identificar qué dependencias tienen sueldos netos medios significativamente diferentes entre sí, realizamos un **análisis post-hoc** utilizando la **prueba de Tukey**. Esta prueba realiza múltiples comparaciones, ajustando automáticamente el nivel de significancia para determinar qué pares de grupos difieren significativamente.

El análisis de Tukey muestra que la dependencia del conjunto que presenta mayores diferencias es la **Dirección de Servicios Generales**:

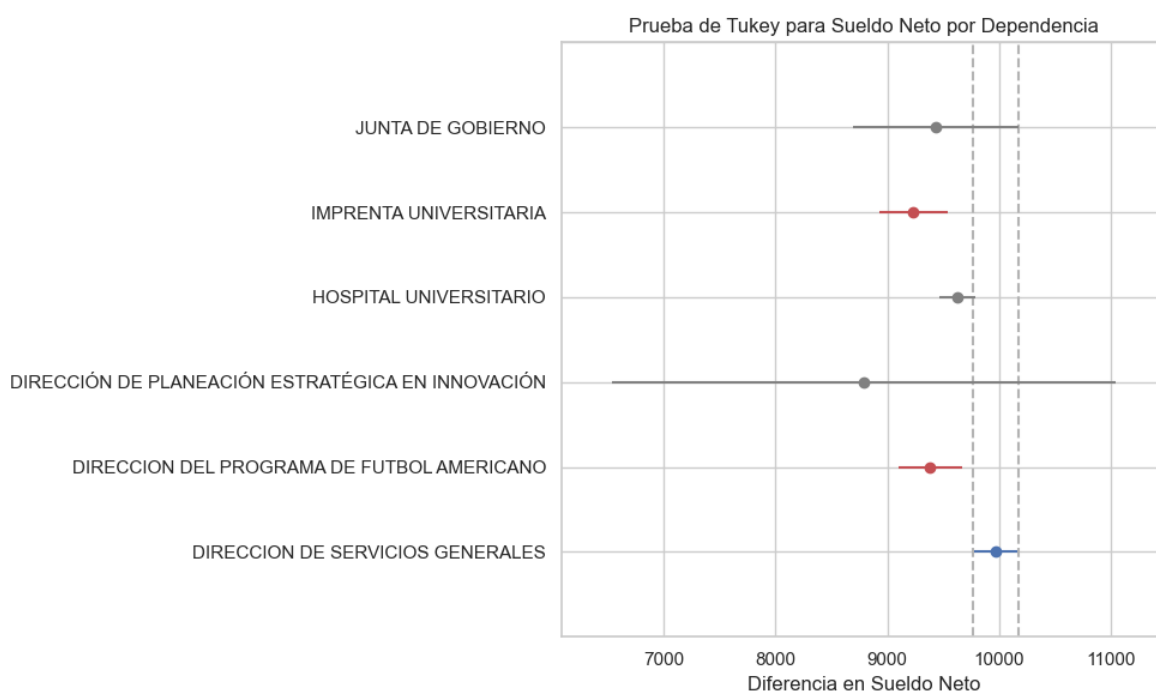


Ilustración 13: Análisis de Tukey para la Dirección de Servicios Generales.

Observamos que el sueldo medio de esta dependencia difiere significativamente de los sueldos medios de la Junta de Gobierno, la Imprenta Universitaria y la Dirección del Programa de Fútbol Americano.

Conclusión

En esta práctica de análisis exploratorio, se ha realizado un estudio exhaustivo que revela hallazgos significativos sobre la distribución de sueldos en las dependencias pertenecientes a la universidad.

Se analizaron 636,201 registros de empleados, en los cuales se detectó una alta variabilidad en los sueldos, tanto a nivel general (desviación típica general de alrededor de \$9,500 MXN) como entre dependencias, con una diferencia de aproximadamente \$40,000 MXN entre los sueldos promedio de las dependencias con más alto y más bajo sueldo, respectivamente. También se observó alta variabilidad dentro de las mismas dependencias, con la Rectoría alcanzando una desviación típica interna de aprox. \$27,000 MXN, y entre tipos de empleo.

En el análisis estadístico, las pruebas de normalidad indicaron que los sueldos netos no siguen una distribución normal en ninguna dependencia, lo cual justificó el uso de pruebas no paramétricas como Kruskal-Wallis y el análisis post-hoc de Tukey.

La prueba de Kruskal-Wallis reveló diferencias significativas entre las medianas de sueldos netos de las 6 dependencias con menor sueldo promedio, destacando la Dirección de Servicios Generales como la más dispar de las dependencias analizadas.

En conclusión, este análisis proporciona una base sólida para futuras investigaciones sobre las políticas salariales en la UANL y sirve como base para investigaciones más específicas sobre las tendencias salariales dentro de la Universidad y su impacto en la vida de los trabajadores